QA
76
M94
SI Rep.

M U D P I E no. 18

Museum and University Data, Program and Information Exchange

(PLEASE NOTE: This issue of MUDPIE is going to all participants in the 1969 Cambridge Symposium-see page 2-who are not already on the list. If this includes you, and you wish to receive later issues, please send your name and address to MUDPIE, Division of Reptiles and Amphibians, National Museum of Natural History, Washington, DC 20560.)

AVAILABLE PROGRAMS

The Food and Agriculture Organization of the United Nations has issued a set of computer programs for fish stock assessment, written by various programmers and compiled for FAO by Norman J. Abramson. The set is identified as FIRD/T101, "Computer Programs for Fish Stock Assessment," and can be obtained upon request from the Research Information Section, Fishery Resources Division, FAO, Via delle Terme di Caracalla, Rome 00100, Italy. The programs listed include:

AGECOM--Estimates age composition using double sampling scheme with length as strata. Also provides estimates assuming simple random sampling of aged fish. FORTRAN. Programmed by D.J.Mackett, Cal. Fish & Game.

BGC 2 --Fits the von Bertalanffy growth-in-length curve to equally spaced age groups with unequal sample sizes for separate ages. N.J.Abramson, Cal. Fish & Game. FORTRAN.

BGC 3 --As BGC 2, but for unequally spaced age groups. FORTRAN. P.K.Tomlinson, Cal. Fish & Game.

BGC 4 --Estimates the parameters K and L sub infinity of the von Bertalanffy growth-in-length curve when only the lengths of individual fish at two points in time are known. This allows the curve to be fitted to tag release and recovery data. FORTRAN. P.K.Tomlinson, Cal. Fish & Game.

WTLN --Fits a curve giving weight as a function of length, produces a table of fitted weights and lengths, and provides various related statistics. FORTRAN. N.J.Abramson, Cal. Fish & Game.

NORMSEP-Separates length frequency sampling distributions into component normal distributions. Used to estimate age group relative abundance in length samples of unageable species. V.Hasselblad, modif. by P.K.Tomlinson. FORTRAN.

An additional nine programs are given, dealing rather closely with commercial fisheries problems such as estimation of fishing power, mortality rates, yield curves, and spawner-recruit curves. --JAP.

SMITHSONIAN
JUL 15 1971
LIBRARIES

DATA PROCESSING IN BIOLOGY AND GEOLOGY

A book with the title above has just been published by the Academic Press in cooperation with the Systematics Association of England, as their special volume no. 3. It is edited by J. L. Cutbill, and includes the proceedings of a symposium held at the University of Cambridge, 24-26 Sept., 1969. As is invariably true of symposium volumes, this one contains its share of pot-boilers that probably should not have been presented and certainly should not have been published. It is still a a very useful summary of the state of the art in the application of computers to museum work, however, and I think everyone who reads MUDPIE should also read this volume. I had thought of listing the titles and summing up some of the work here, but I decided against this, feeling that it would be wiser to persuade all of you to examine the book critically. At least in part this is because there were 135 people who attended the symposium, and practically none of them are people who currently receive MUDPIE. There clearly exist at least two large groups of people interested in the same thing--the handling of data and information in museums--and who knows how many more such groups there may be? The painful consequence of this, of course, is the continued proliferation of duplicative effort remarked upon in MUDPIE no. 3, way back in January, 1968. We don't seem to have come too far in coordinating our efforts since then, MUDPIE not withstanding.--JAP.

KEEPING UP WITH ARPA-I

In MUDPIE no. 14 there is a short introduction to ARPA--the Advanced Research Projects Agency of the Department of Defense--and its Computer Network. I remarked at that time that it will not hurt to keep up with their activities, even though museums cannot afford their high-priced action. This is the first installment of "keeping up."

A news story in Datamation, May, 1971, says that ARPA Computer Network plans to have access by 1974 to an on-line information storage and retrieval system with a trillion-bit memory. The system will be located with and on-line to the ILLIAC IV at NASA in Ames, California. Computer Corporation of America, in Cambridge, Massachusetts, will begin immediately to assemble a facility to store large files at the minimum expense, to provide a community of users with file-sharing ability, a uniform method of accessing remote files, and advanced data management services.

The system will store such files as a 100 billion-bit data base of historical meteorological information. It will have little or no computational ability, because it will be set up to pass on data to the user's computer for manipulation. Eventually, the data will be passed on to the ILLIAC IV over high-speed lines, if the amount of computation is too much for a local machine.--JAP.

AN INFORMATION SYSTEM FOR THE GEOSCIENCES

The 10 member Committee on Geoscience Information of the American Geological Institute has prepared an official statement approved by the 17 member societies of the AGI, which is to serve as a base for continued planning. We have abstracted below some interesting sections from the official text of the statement.--JAP

DEFINITION: We define an "information system" as a functioning program for the efficient transfer of information, involving all conventional channels and services, updated to provide for the explosive growth of geoscience information during the past several decades, and employing new techniques introduced with the development of electronic data processing.

SCOPE: The system should encompass all types of geoscience information. It should be concerned with the following, and other areas of information transfer:

1. Formal meetings and symposia.
2. Primary journals and monographs.
3. Translations and review journals.
4. Bibliographies, abstract journals, and indexes.
5. Archival holdings of libraries.
6. Guidebooks and informal reports.
7. Theses and dissertations.
8. Data collections.
9. Glossaries and thesauri.
10. Maps, charts and photographs.
11. Collections of interest to geoscientists (type specimens, cores, well logs, thin sections, etc.).

CHARACTERISTICS: The system should be operated for geoscientists by geoscientists. The system must not be static. It must be flexible enough to accept change when indicated, to admit new services that appear to be required, and to abandon services that are unused for whatever reasons. System planning should be centralized for economy and efficiency, but system operations should be decentralized in order to utilize the experience and capabilities of those organizations that are now engaged in the dissemination of geoscience information, or who may desire to become so engaged in the future. The system should include the information services provided by all types of organizations within the geoscience community: societies, universities, state and federal agencies, industry, and commercial information companies. The system should operate economically. The system should be developed with cognizance of similar systems in other scientific disciplines, in order to promote efficiency and economy. It should overlap those systems to include the subject matter of other disciplines that may be of special interest to geoscientists. The system should be user oriented.

(The report concludes with a list of RESULTS, most of which are obvious from the above.)

INFORMATION SYSTEMS SURVEY
"PRC"
(Plant Records Center)

The Plant Records Center (PRC) is an organization aimed at applying computer techniques to the control and processing of information about the cultivated plants of North America. It is located at the Tyler Arboretum in Lima, Pennsylvania. Dr. Robert D. MacDonald is Director. He has a staff of approximately seven people including a botanist/taxonomist, a botanist/ systems analyst, an editor, an administrative assistant and four people involved in data automation. Development of computer programs and computer processing is performed under contract by a service bureau.

The Center is operating under a two-year grant from Longwood Gardens Foundation, and hopes to receive subsequent grants for a total period of ten years during which time it will complete development of a major data base concerning living plant collections of North America. After that they plan to be selfsupporting by offering for sale a variety of products in the form of computer reports. They are already selling some products and services on a limited scale as their data base evolves.

MAJOR FUNCTIONS--PRC is currently performing three major functions. First, they are developing a machine readable data base inventorying the living plant collections of botanical gardens and arboreta and of miscellaneous other special collections in North America. This project will be discussed in more detail in the remainder of this report.

Second, they are preparing registrations and checklists of cultivar names for selected genera by genus to the infraspecific level. These lists, in addition to the scientific name, provide a bibliographic reference to place of publication and, in some cases, the place of origin of the cultivar and a brief note on its distinguishing characteristics. Validity of data is checked by the registration authority which is normally part of a particular plant society. PRC is working on their first check list now. In addition to the above information, they will develop distribution maps based on the existence of specimens in botanical gardens.

Their third function is to provide a plant directory service or plant buyers guide to rare and unusual plants.

INVENTORY OF LIVING PLANT COLLECTIONS-- The living plant inventory is now under development. The inventory consists of a growing computer-readable file of records for plant specimens in botanical gardens, arboreta and other special sites in North America. It is being built by inventorying one site at a time and generating a computer record for each living plant specimen at the site. At the present time data have been input to the system for Longwood Gardens and the Fairchild Arboretum. Data for the US National Arboretum are now being converted to computer records.

Fifteen other gardens are preparing to enter their records into the system. The expectation is that at the end of ten years, over 150 collections covering one million plant records will be in the system.

Original data are gotten from the garden or arboretum by microfilming their existing accession records which are usually in the form of card files. A team from PRC visits the garden, sequence numbers their cards and microfilms them using a portable microfilming device which produces cartridges. In preparing data for conversion to machine-readable form, family names are coded using Gould's system although the plain text family name is now also carried in the record. The data are converted to machine-readable form using a keypunch or a Honeywell keytape system. All data are captured in upper case. The operator works directly from the microfilm using a microfilm reader. She skips information which she cannot decipher and makes notes on botanical or editorial help which she will need to complete the record.

After the records have been converted to machine readable form, a card image listing is made and edited at PRC. A listing, including all corrections made at PRC, is returned to the garden for final review. Their corrections are made and the data are added to the data bank. PRC makes no taxonomic decisions on the data received. They do check names for proper form. A wide variety of reports can be prepared from the data bank.

Forms have been developed which participating gardens will use in the future to maintain their own accession files and to transmit new records to PRC. New records and corrections and deletions of existing records will be entered into the system in a batch periodically.--Harriet R. Meadow.

RECENT LITERATURE

Biss, K., R. Chien, & F. Stahl. R2--A natural language question-answering system. AFIPS Conference Proceedings, vol. 38, 1971: 303-308.

David, E. E. Computing from the Communication Point of View. Advances in Computers, vol. 10, 1970: 109-128.

Eddy, A. A meteorologically oriented computer model of an ecosystem. Transactions New York Academy of Sciences, vol. 31, 1969: 618-628.

Emshoff, J. R. & R. L. Sisson. Design and Use of Computer Simulation Models. MacMillan, New York, 1970: xvii +302. (Devoted almost entirely to simulation of business problems. Practically no use to MUDPIErs.)

Grenander, U. A Unified Approach to Pattern Analysis. Advances in Computers, vol. 10, 1970: 175-216. (The author devotes a section to taxonomic practices and another to diagnostic determinations, erring slightly when he compares a diagnostic flow sequence to a phylogenetic tree. The paper is worth the attention of taxonomists.--JAP)

Hall, T. W. Implementation of an interactive conference system. AFIPS Conference Proceedings, vol. 38, 1971: 217-229. (A description of a method for holding business meetings or conferences without leaving your office, using a time-shared computer. The computer permits motions, discussion, and votes, and then prepares a final record of the conference!)

Hare, V. C. BASIC Programming. Harcourt, Brace & World, New York, 1970: x + 246. (Another substitute for the BASIC manual handed out by time-share contractors--but this one is good enough to buy.)

Holloway, J. D. The biogeographical analysis of a transect sample of the moth fauna of Mt. Kinabalu, Sabah, using numerical methods. Biological Journal of the Linnean Society, vol. 2, 1970: 259-286.

Jacobs, W. W. A structure for systems that plan abstractly. AFIPS Conference Proceedings, vol. 38, 1971: 357-364.

Ledley, R. S. Use of Computers in Biomedical Pattern Recognition. Advances in Computers, vol. 10, 1970: 217-252.

Lewis, W. H. An ecological vocabulary for the Flora North America program. FNA Report 57, 1971: 1-23.

Meadow, C. T. Man-Machine Communication. Wiley-Interscience, New York, 1970: xix + 422. (An easy to read summary of all kinds of interaction between computer and man. Highly recommended for the novice who wants to know what is going on.)

Prosser, F., & D. D. Jensen. Computer generated repeatable tests. AFIPS Conference Proceedings, vol. 38, 1971: 295-301.

Shetler, S. G. Flora North America as an information system. BioScience, vol. 21, 1971: 524-532. (This issue of BioScience includes a questionnaire to be filled out by systematic botanists interested in participating in the FNA project. Since there are broad plans for continued automation of the basic data, as discussed in this article by Shetler, botanically inclined MUDPIE readers should obtain a copy of the questionnaire and send it in.)

Tryon, R. C. & D. E. Bailey. Cluster Analysis. McGraw-Hill, New York, 1970: xviii + 347 pp. (An in-depth study of cluster analysis, with directions on the use of a series of chained programs in FORTRAN IV, which are available for use as directed here. The authors are primarily concerned with behavior analysis, but there is some discussion of the problems of taxonomic use of the techniques.--JAP)

Tymes, L. R. TYMNET--A terminal-oriented communication network. AFIPS Conference Proceedings, vol. 38, 1971: 211-216.

van Dam, A., & D. E. Rice. Computers and Publishing: Writing, Editing, and Printing. Advances in Computers, vol. 10, 1970: 145-174.

Wied, G. L. & G. F. Bahr (Editors). Automated Cell Identification and Cell Sorting. Academic Press, New York, 1970: xi + 403. (Most of the included papers are concerned with biochemical and physical techniques in preparing cells for examination by automated techniques. One paper, however, is a very informative and useful summary of a group of computer programs called TICAS, for Taxonomic Intra-Cellular Analytic System. This paper is important to any taxonomist interested in the possibilities of computer use in the organization of taxonomic information, and who thinks that numerical taxonomy is not the entire answer to the problems of systematics. See the remarks about n. t. on p. 346 of the paper, which is: Wied, G. L., G. F. Bahr, & P. H. Bartels; Automatic Analysis of Cell Images by TICAS; pp. 195-360.--JAP)

July, 1971
Division of Reptiles and Amphibians
National Museum of Natural History
Washington, DC 20560